Mon. Not. R. Astron. Soc. **269**, 97–104 (1994)

# Automated classification of stellar spectra – I. Initial results with artificial neural networks

T. von Hippel,[1] L.J. Storrie-Lombardi,[1] M.C. Storrie-Lombardi[1] and M.J. Irwin[2]
[1] *Institute of Astronomy, Madingley Road, Cambridge CB3 0HA*
[2] *Royal Greenwich Observatory, Madingley Road, Cambridge CB3 0EZ*

Accepted 1994 February 7. Received 1994 February 3; in original form 1993 July 22

**ABSTRACT**

We have initiated a project to classify stellar spectra automatically from high-dispersion objective prism plates. The automated technique presented here is a simple back-propagation neural network, and is based on the visual classification work of Houk. The plate material (Houk's) is currently being digitized, and contains $\approx 10^5$ stars down to $V \approx 11$ at $\approx$ 2-Å resolution from $\approx$ 3850 to 5150 Å. For this first paper in the series we report on the results of 575 stars digitized from 6 plates. We find that even with the limited data set now in hand we can determine the temperature classification to better than 1.7 spectral subtypes from B3 to M4. Our current sample size provides insufficient training set material to generate luminosity and metallicity classifications. Our eventual aims in this project are (1) to create a large and homogeneous digital stellar spectral library; (2) to create a well-understood and robust automatic classification algorithm which can determine temperatures, luminosities and metallicities for a wide variety of spectral types; (3) to use these data, supplemented by deeper plate material, for the study of Galactic structure and chemical evolution; and (4) to find unusual or new classes of objects.

**Key words:** methods: data analysis – methods: numerical – stars: fundamental parameters – Galaxy: stellar content.

## 1 INTRODUCTION

The classification of stellar spectra (e.g. Cannon & Pickering 1918-24; Morgan, Keenan & Kellman 1943; Johnson & Morgan 1953) has long been a fundamental tool for determining the physical parameters of bright stars. While early work concentrated on stellar temperatures and then luminosities, it wasn't long before a significant number of stars were found with unusually low abundances, abundance anomalies, and unusual combinations of stellar types. The distribution of stellar temperatures and luminosities led Hertzsprung (1911) and later Russell (1914) to plot stars in the now-famous HR diagram. The distribution of combinations of stellar types aided early explorations of binary creation scenarios (e.g. Aitken 1935) and enhanced the early work on the determination of stellar masses from spectroscopic binaries (e.g. Russell 1912). The realization that many of the low-abundance stars had high relative velocities (Roman 1954) led to the separation of Galactic stars into Populations I and II, and to the formation scenario for the Galaxy of Eggen, Lynden-Bell & Sandage (1962). Spectral classification has continued to play an active role in the identification of both normal and unusual stellar types. The distribution of normal stars, corrected for the volume element sampled, leads to the present-day mass function, and thereby to the initial mass function, providing a major clue to the outcome of local star formation. The local stellar density yields the mass of a major component of the Galaxy, while the decrease in the density with distance from the Galactic plane indicates the diffusion with age of a given sub-population.

Over the last 20 years, Houk and co-workers (Houk & Cowley 1975; Houk 1978, 1982; Houk & Smith-Moore 1988) have been creating one of the greatest modern stellar classification libraries by determining the two-dimensional (and often three-dimensional) spectral types of all stars in the Henry Draper Catalogue (Cannon & Pickering 1918-24, hereafter HD). The four catalogues published to date by Houk include over 120 000 stars covering the full range of temperatures and luminosities, and a wide range of metallicities down to the HD magnitude limit of $\approx$ 11. Errors in human classification are hard to quantify, but Houk reclassifies $\approx$ 10 per cent of her stars without knowledge of her earlier classification, and finds an rms error of $\approx$ 0.6 spectral subtypes, corresponding to an uncertainty of $\approx$ 2000 K for a 20 000-K star ($\approx$ B3) or $\approx$ 100 K for a 5000-K star ($\approx$ K0, Mihalas & Binney 1981). In the luminosity class domain, Houk's rms error is $\approx$ 0.25 luminosity types. Given the importance and the time-consuming nature

of human visual classification of stellar spectra, considerable effort has been made towards automating the process (see Kurtz 1983, and references therein). Some successes have been achieved, though no readily usable and robust techniques have emerged to date. Besides a hoped for increase in classification speed, automation should also make it easier to quantify random and systematic errors, and to study the effects of lower resolution and lower signal-to-noise ratio on classification errors. Automated techniques may not do as well as a well-trained human classifier, especially for the most difficult or unusual types, and it is expected that a human classifier will have to study objects that pose particular difficulties for the automated techniques. On the other hand, an automated, objective and repeatable technique will provide a basis for more detailed insights into the astrophysics underpinning the stellar spectral features.

Given the tremendous amount of high-quality spectral classifications done by Houk and her extensive and homogenous plate material, we seek to build an artificial expert system for stellar spectral classification based on Houk as our expert. Ideally, one might want the astrophysical parameters of temperature, surface gravity and abundances to come from automated techniques, rather than classifications. While we agree in principle, we note that at the present time automation based on classifications should be superior, since (1) not enough stars have their astrophysical parameters precisely defined to build the type of bulk classification system we envision; (2) classifications can be mapped into the astrophysical parameters with well-studied relations (e.g. Popper 1980, and references therein); (3) this mapping can be repeated whenever more up-to-date and accurate relations are determined; and (4) classifications are considered inviolate, since they are based on a series of standards, and therefore should not change with time. We note that recent advances in plate measuring (e.g. the Automatic Plate Measuring Facility in Cambridge, Kibblewhite et al. 1984, hereafter APM) and the steady advances in computer speed are also key ingredients in our automation efforts.

In this paper, the first in a series, we examine an expert system based on a back-propagation artificial neural network (hereafter ANN). Other techniques can be readily envisioned, for example classical template matching in the form of line fitting (Jones 1966), cross-correlation and minimum vector distance (Kurtz 1983), to name a few. Line fitting techniques suffer from the disadvantage that one has first to know the approximate stellar type before determining which lines to fit, otherwise very different features will be found at the same wavelengths. Cross-correlation in its simplest form also weights comparisons towards the strongest lines, which are not necessarily the features with the highest weight in classification determination. Cross-correlation also requires a well-populated library of homogenous quality, but this is something which we could eventually create. Minimum vector distance techniques (Kurtz 1983) have had some successes, and also require good libraries. All these classical techniques are based on linear operations. Since we might expect to find rather subtle non-linear relationships between the temperature, surface gravity and metallicity line indicators, a classification scheme that copes with non-linear relationships between parameters ought to offer significant advantages. Supervised ANNs can not only solve non-linear optimization problems, but also have obvious similarities to a human classifier. Furthermore, with a supervised ANN the designers of the expert system do not have to be classification experts themselves, but rather can let the answers of the original expert train the ANN. Such ANNs also seem to be robust to stellar libraries of heterogenous quality, and they can be designed (Richard & Lippmann 1991) to provide not only classification answers, but also uncertainties associated with their answers.

**Table 1.** Plates scanned.

| date | plate | RA | Dec | $l$ | $b$ | num |
|---|---|---|---|---|---|---|
| 27/ 5/68 | 2280 | 4:13.2 | -1:46 | 194 | -35 | 76 |
| 28/ 5/68 | 2303 | 3:28.6 | 1:23 | 182 | -43 | 99 |
| 16/12/68 | 3316 | 20:26.9 | -0:44 | 44 | -21 | 96 |
| 12/ 1/69 | 3504 | 5:29.7 | 4:43 | 199 | -16 | 136 |
| 15/ 7/69 | 4825 | 23:14.6 | -4:45 | 73 | -58 | 73 |
| 9/ 8/74 | 16506 | 4:35.6 | 2:58 | 193 | -28 | 95 |

## 2 DATA REDUCTION AND PRE-PROCESSING

The data reduction for this project is essentially the conversion of Houk's objective prism plates and her catalogue into one-dimensional digital spectra tagged with her spectral types and luminosity classes. The raw data are $\approx$ 1000 IIaO high-dispersion (108 Å mm$^{-1}$ at H$\gamma$) objective prism plates taken with the Michigan Curtis Schmidt telescope at Cerro Tololo, Chile. Each plate is 20 $\times$ 20 cm$^2$ in size, and at a plate scale of 96.6 arcsec mm$^{-1}$ subtends an area of 5° $\times$ 5° on the sky. We report here the results from the scanning of 6 plates containing 575 HD stars which Houk classified and assigned a quality of 1 or 2 (on a 1 to 4 system). Of these 575 stars, 371 are quality 1 (65 per cent) and 204 are quality 2 (35 per cent). There are 138 quality 3 stars and 11 quality 4 stars on these 6 plates. Details of the plate material can be found in Table 1, where we list the date the plate was taken, Houk's plate code, the J2000.0 equatorial coordinates and Galactic coordinates of the plate centres, and the number of non-overlapped quality 1 and 2 spectra extracted from each plate.

The APM facility in Cambridge was used to measure the plates. General details of this facility are given in Kibblewhite et al. (1984), whilst the paper by Hewett et al. (1985) describes the general use of this system for measuring objective prism plates. The scanning procedure uses the *HST* Guide Star Catalog as a position driver for the scanner, then extracts digitized scans 18 $\times$ 2 mm$^2$ in area around each widened spectrum at 22.8 $\mu$m pixel$^{-1}$ resolution (corresponding to a 768 $\times$ 80 array). After finding and subtracting the general sky background level, the digitized scans are summed orthogonally to the dispersion axis to create one-dimensional, pixel versus density spectra. Following this, the extracted spectra are matched against the Houk catalogues, and for the moment only those for which Houk has determined a spectral type are kept. In future work we plan to keep all spectra with sufficient signal-to-noise ratio for classification, but at present we are just building the expert system, and not classifying previously unclassified stars.

The next data reduction step was differential wavelength calibration of all the spectra within each plate. An objective prism introduces a quadratic distortion of the position of a spectrum on the prism plate compared to a direct plate (in this case the *HST* Guide Star Catalog), thereby causing

the spectrum to shift slightly (up to several pixels) from its expected position. This distortion was mapped as a function of the standard coordinates of the objects by cross-correlating each high-quality spectrum with a high-S/N template star of approximately the same type. A quadratic transformation relating direct and prism coordinates was then derived and the solution applied to all stars regardless of S/N. This places all the stars on the same (arbitrary) wavelength system.

The final stage in the pre-processing was to define a stellar continuum so that the information in the spectra could be partitioned into line and continuum features. We did this because we wanted to examine classification schemes based on (1) the complete spectra, (2) line information alone, and (3) continuum information alone. The line-only case is the generic method employed by human classifiers, whereas the continuum-only method is equivalent to classification based on photometric colours. The advantages and problems of each of these three cases will be discussed in greater detail below (see Section 4). The continuum estimation was accomplished by first running a 51-pixel median filter (Tukey 1971) over each spectrum, followed by a 25-pixel boxcar filter. All absorption (or emission) features with scale length less than $\approx$ 25 pixels are removed by this processing. This is sufficient to track over all the main absorption lines, though it underestimates the continuum in areas where the spectral shape changes rapidly, especially through the major molecular features. This will have no effect on our analysis of the line-only and continuum-only cases, as long as the continuum fits are consistent for a given spectral type and no residual continuum slope remains. A greater quantity of data, including significantly reddened stars, will be required to test the continuum fits fully. The line-only cases were created by dividing the spectra by their continua.

As samples of the spectra used in this work, Fig. 1 presents spectra of 6 stars of various spectral types, along with their continuum fits. From top to bottom the stars are classified as B9 V, A2 V, F3 V, G3 V, K5 III and M2 III, respectively. The horizontal axis is wavelength in Å, and the vertical axis is emulsion density in arbitrary units. The spectra have S/N $\gg$ 100 and an effective resolution of $\approx$ 2 – 3 Å, resulting in 382 resolution elements over the useful range of 3843 – 5150 Å. Note that for the later spectral types the more limited flux in the blue increases the noise markedly shortward of 4000 Å.

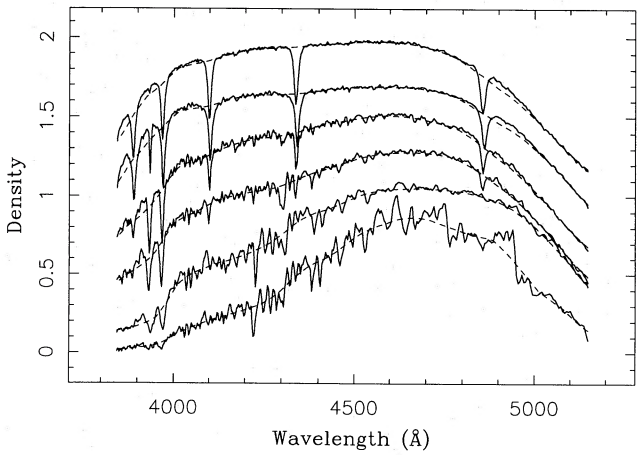


**Figure 1.** Sample spectra plotted with flux in arbitrary units.

## 3 CLASSIFICATION ALGORITHM

### 3.1 Background

For this study we have utilized an artificial intelligence technique known as artificial neural networks (ANNs) for stellar classification. Originally derived from simplified models of human central nervous system activity (McCullogh & Pitts 1943; Hopfield 1986), ANNs have found utility in signal processing (Widrow & Winter 1988), adaptive optics (Merkle 1988; Sandler et al. 1991; Wizinowich et al. 1991), star/galaxy distinction (Odewahn et al. 1991), and galaxy classification (Storrie-Lombardi et al. 1992).

Discovery of an error minimization algorithm for training multiple layers of ANN nodes (Werbos 1974; Parker 1985; Rumelhart, Hinton & Williams 1986) has produced considerable interest in a very simple gradient descent, non-linear minimization ANN technique known as back-propagation. Back-propagation ANNs consist of a series of nodes arranged in layers and analogous to biological neurons with inputs (dendrites), summation nodes (cell bodies), outputs (axons), and weighted connections between layers (synapses). The classical back-propagation algorithm requires a minimum of three layers: an input layer for bringing the data into the ANN, at least one hidden layer to effect a non-linear mapping from input to output space, and an output layer to provide classification. We have used 383 input nodes, one for each of the 382 spectral resolution elements and one as a bias node which has a constant value (see Fig. 2). Weights connecting the bias to the hidden layer and output nodes are modified during training exactly like all other weights. We term any layer between the input and output layers as a hidden layer. We have confined ourselves to the case of a single hidden layer and a single output node for the purposes of this study to minimize the number of free parameters in our system. We have one output node to provide the temperature classification. (We note here that we use the word classification throughout this paper as it is used in the stellar spectroscopy community, i.e. to mean the position of a star in the HR diagram.)

In the standard back-propagation algorithm, the output is defined as a vector $\boldsymbol{o}$ in the classification space with an individual node assigned for each class. In our case this would produce a 59-dimensional output vector if we considered each temperature classification as a discrete class. The ANN starts with a random initial weight state and guesses the identity of an incoming target vector by producing an output vector with the value of each output node lying somewhere between 0 and 1. For each star, it then compares its output to the desired classification vector $\boldsymbol{d}$ generated by the human expert. For example, we might define $\boldsymbol{d} = (1,0,0,0,...)$ for class B0, $\boldsymbol{d} = (0,1,0,0,0)$ for class B1, etc.

The comparison is done in terms of a cost function, usually of the form

$$E = \frac{1}{2}\sum_k (o_k - d_k)^2$$

summed over the vector components. Because during training the ANN adjusts its weight space according to the decisions of a human expert, we term it a supervised network. This cost function is minimized with respect to the free parameters,

the weights $w_{ij}$, connecting each layer. Minimization over the weights $w_{ij}$ is done using the chain rule (gradient descent). The weights are updated *backwards* from the output layer through the hidden layer(s), and the rate of updating is controlled by the learning coefficient, $\eta$. A second coefficient, $\alpha$ (often called the momentum), can damp the oscillations in minimization sometimes seen as the ANN searches through weight space for an optimal solution. Training set size determines the optimal values for $\eta$ and $\alpha$ (Eaton & Olivier 1992). (Note that although $\eta$ and $\alpha$ are the most commonly used symbols for the learning and momentum coefficients, respectively, other symbols do appear in the literature. Some authors have even reversed the usage of $\eta$ and $\alpha$.)

At each node at layer $s$ of the ANN, a linear combination over the input $x_i$ from the previous layer $s-1$ is given by $I_j^{(s)} = \sum_i w_{ij}^{(s)} x_i^{(s-1)}$. The node fires according to a non-linear sigmoid threshold function usually of the form $f(z) = 1/(1+\exp(-z))$ (in the interval [0,1]) or $f(z) = \tanh(z)$ (in the interval [−1,1]). The ANN functions as a non-linear operator transforming the representation of objects in the input parameter space to the classification space.

After completion of the learning process (i.e. optimizing the weights $w_{ij}$) according to a given training set, the ANN is ready to analyse new data. As we present each input vector the ANN produces an output vector containing its classification coding. In the one output per class configuration, the output of an ANN provides a Bayesian a posteriori probability for class membership (e.g. Richard & Lippmann 1991; Storrie-Lombardi et al. 1992).

A single output ANN does not have the ability to attach a classification probability to each object, but it provides another interesting capability. Back-propagation ANNs given analogue inputs can generate an analogue output that replicates any continuous, differentiable function (Cardaliaguet & Euvrard 1992), and even chaotic systems provide excellent signals for ANN mapping (Mpitsos & Burton 1992). In such a system, the hidden nodes continue to use a non-linear algorithm, but the output layer produces a linear output. It can even predict two functions at the same time (e.g. temperature and luminosity). The single output node back-propagation neural network can function as a non-linear least-squares optimization algorithm producing a non-linear fit to a continuous classification function. In our case it transforms the stellar spectral bins into a continuous temperature classification.

### 3.2 ANN design

ANN design depends on finding an optimally efficient network architecture for a particular data set. The principal factors include size of the training set, number of input parameters, number of hidden layer nodes, and number of output nodes. We have maximized the size of the training set to approximately 500 objects and yet utilized all 575 spectra as test objects by setting up a revolving train/test cycle. The revolving train/test cycle procedure further allows us to test for any possible plate-to-plate systematic effects, while still maintaining for each train/test combination an unseen testing set. The entire data base is never shown to a given ANN, but rather five plates of stellar spectra are used to train the ANN and one plate of spectra to test the ANN.

For our input layer we use 382 nodes (i.e. the number of spectral resolution elements) for all of our ANNs. If we were to use the 59 temperature classifications recognized by Houk as discrete classes, a complete network with one output node per class and three nodes in the hidden layer would require a total of 177 weights to connect the hidden and output layers fully. While the probabilistic output associated with class membership available to us in the one output per class format provides numerous advantages for such activities as developing reliability estimates for the classifications, the number of additional weights involved in creating individual output nodes brings considerable overheads (e.g. increased CPU time) and complexity to the task of stellar spectral classification. The complexity of the ANN depends on the number of inputs, hidden nodes, layers, outputs, and connections. Increased complexity can certainly produce an ANN capable of a much richer and more precise set of decision boundaries for classification. In fact, as the number of free parameters approaches the number of objects in the training set, the ANN can often completely memorize the training set and produce a 100 per cent accurate classification of objects when tested on its original training set. At first glance such memorization might appear reassuring and seem to imply that the parameters presented to the ANN do have an underlying pattern capable of predicting class membership. Unfortunately, such an ANN usually performs more like a simple look-up table and may have increasing difficulty generalizing its classification algorithm to new test data. Since a neural network is a non-linear least-squares optimization algorithm, the convergence properties are determined by the ratio of data input points to free parameters (i.e. weights).

A decrease in the total number of weights can be accomplished by only partially connecting nodes. However, the problem of optimal connectivity remains a subject of considerable research effort. For the purpose of the present study we have chosen to leave our network fully connected and decrease the total number of weights by decreasing the number of nodes in the hidden and output layers. We have constructed a very simple, minimalist ANN. Back-propagation networks using a single hidden layer train more rapidly and fall into inefficient local minima less frequently than ANNs with two hidden layers (de Villiers & Barnard 1993). We utilize only a single hidden layer containing three nodes. In tests on hidden layer size we have investigated three, five, and seven nodes. The five- and seven-node cases have the same random errors as the three-node case over 1000 training cycles. The five-node case has 3 per cent greater systematic errors than the three-node case, and the seven-node case has 38 per cent greater systematic errors.

Finally, we have implemented a single output node for our classification strategy, since temperature classification is a continuous function rather than a discrete set of output classes. The different spectral type subclasses were converted to floating point numbers from 0 (O3) to 1 (M9), following the 59 subclasses recognized by Houk. This simple linearization of the Houk spectral types mimics the human ability to recognize the effects of small changes in temperature. We have investigated both linear and non-linear outputs and find no difference in the output for our particular problem. Since the linear output networks train $\approx$ 3 times more rapidly, we have reported results for the linear output case.

At this time we have confined our classification to the temperature domain, since we do not have a significant number

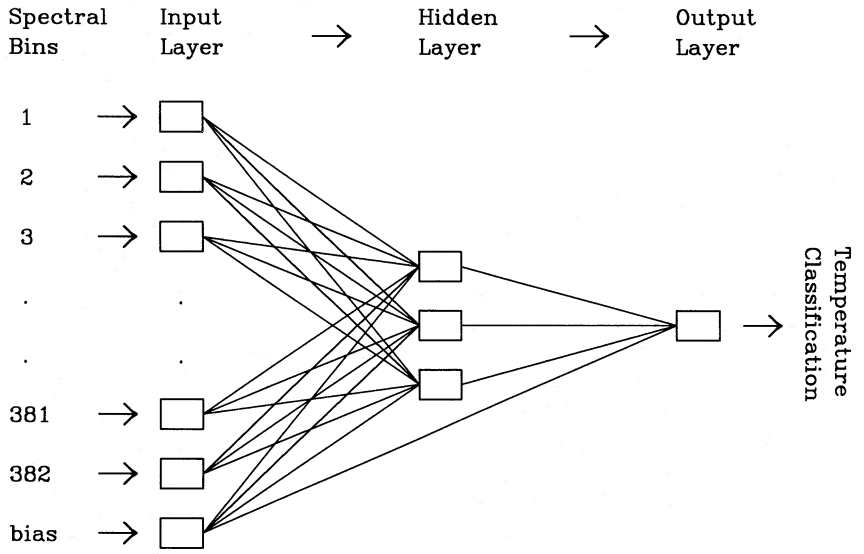


**Figure 2.** ANN architecture.

of stars of different luminosity class at a given temperature class. The spectra described in this series of experiments do not all reside in one luminosity class, but predominantly come from dwarfs at the earlier types and giants at the later types. At this point we have no stars at the classification extremes, either earlier than B3 or later than M4.

Fig. 2 depicts our fully connected (each node connects to every node in the next layer) back-propagation network. We use 382; 3, 1 as the descriptor for the ANNs with 382 inputs (one input node for every spectral resolution element), three hidden nodes and one output node.

### 3.3 Training

Once the basic ANN architecture has been implemented, values for the learning ($\eta$) and momentum ($\alpha$) coefficients must be fixed. Once again these will probably vary as a function of data set size. Future work will deal more exhaustively with the optimum values of momentum, learning, number of passes through the training data and number of hidden nodes. We have found that for our current data set $\alpha = 0.9$ and $\eta = 0.01$ produce stable temperature classifications.

To train, an ANN must evaluate each input vector hundreds or even thousands of times. The major portion of the error minimization occurs in the first few dozen iterations, followed by an exponential decrease in the rate of learning. Reasonable criteria for stopping learning include (1) the rate of learning has effectively approached zero, and (2) a comparison of ANN performance against its training set versus the test set indicates good generalization to the test data without memorization of the training set. Interestingly, back-propagation ANNs using a non-linear hidden layer and linear output layer generally reach maximal performance considerably faster than the classical multi-node, non-linear output system. We have found that our networks reach a stable state and show good generalization after 1000 passes through our training samples. Fig. 3 depicts the learning curve for one of our ANNs with its independent training and testing sets.

Increasing discrepancy between training and testing errors

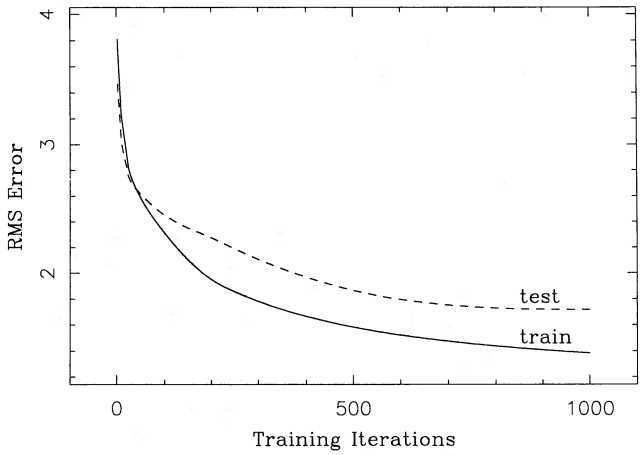


**Figure 3.** Learning curve for a typical ANN. The horizontal axis is the number of training iterations and the vertical axis is the root-mean-square differences of Houk − ANN.

indicates that this ANN has started to reach the limits of its generalization capability. The fact that the testing errors have not passed a minimum means that the ANN was not overtrained, entering into a memorization state.

## 4 RESULTS

In the following discussion we build three separate ANNs, one for complete spectra (continuum plus line information), one for line information only and one for continuum information only. All three ANNs have the same architecture.

In order to maximize the number of stars used to train the ANNs, we trained a given ANN on the spectra from five plates and tested on the spectra from the remaining one plate. We repeated this procedure six times for each possible grouping of plates, thereby cycling through the entire data base for training and testing. Each network begins training from scratch with the same random weights. At present, with 575

spectra containing 382 independent data points (resolution elements) and an ANN architecture with 383 (= 382 input nodes + 1 bias node) × 3 (hidden layer nodes) + 4 (= 3 hidden layer nodes + 1 bias node) weights (free parameters), the final ANN weights are overdetermined by a factor of ≈ 100 (the number of degrees of freedom is reduced by ≈ 2 by the cycling training/testing procedure). This training and testing procedure offered us the advantage of using nearly all 575 stars to train a given ANN, and it allowed us to test whether all plates would give the same results. Plates may not all give the same results, as they vary due to emulsion differences and due to astronomical differences, such as reddening. While it will be shown that the differences between these six plates are small, future work will take a more detailed look at plate emulsion differences, reddening, and other effects. We also see differences in the continuum shapes of stars of the same spectral type on the same plate as a function of magnitude, which is the expected result of the non-linear behaviour of photographic emulsions. While these differences are not important at the present level of precision, they will be calibrated out in future work.

Another persistent problem at present is the large number of objects that are contaminated by overlapping spectra. For the present these objects are found and removed by eye. Undoubtedly some spectra remain that are affected by overlapping objects, and this may degrade the quality of the resulting temperature classifications.

### 4.1 Complete case

For the case where the complete spectra are treated, we normalized the spectra by setting the maximum intensity value to 1, and the minimum to 0. Fig. 4 shows the result of a typical training and testing session for the complete spectra ANNs. Panel (a) is the training case and panel (b) the testing case. The horizontal axis for both panels is the catalogue classification of Houk and the vertical axis is the ANN classification. A line of slope unity is plotted in each panel to allow the eye to asses the quality of the ANN determinations. The self-test plots are created by passing the training set back through the ANN, essentially showing how well the ANN reached the global minimum of the solution space during training. The testing plots are created by passing the previously unseen data through the trained ANN. The testing data will generally have a larger scatter than the self-test data due to the imperfect ability of the ANN to generalize. None the less, in all cases with these data, the test data have only slightly more scatter than the training data, and it can be seen that good quality classifications result.

This can be seen quantitatively in Tables 2 and 3. Table 2 presents the systematic classification offsets and uncertainties for the training and testing data. The first column identifies the testing plate of the ANN. The second and fifth columns list the systematic classification offsets for the training and testing cases, in the sense of Houk value − ANN value, in units of integral spectral subtypes (e.g. 1.0 is the difference between an F6 and an F7 star). The third and sixth columns list the classification uncertainties, which are the root-mean-square differences of Houk − ANN. The fourth and seventh columns list the distance from the mean that contains 68 per cent of the classification differences. If the distribution of classification differences were normally distributed, then $\sigma$ and

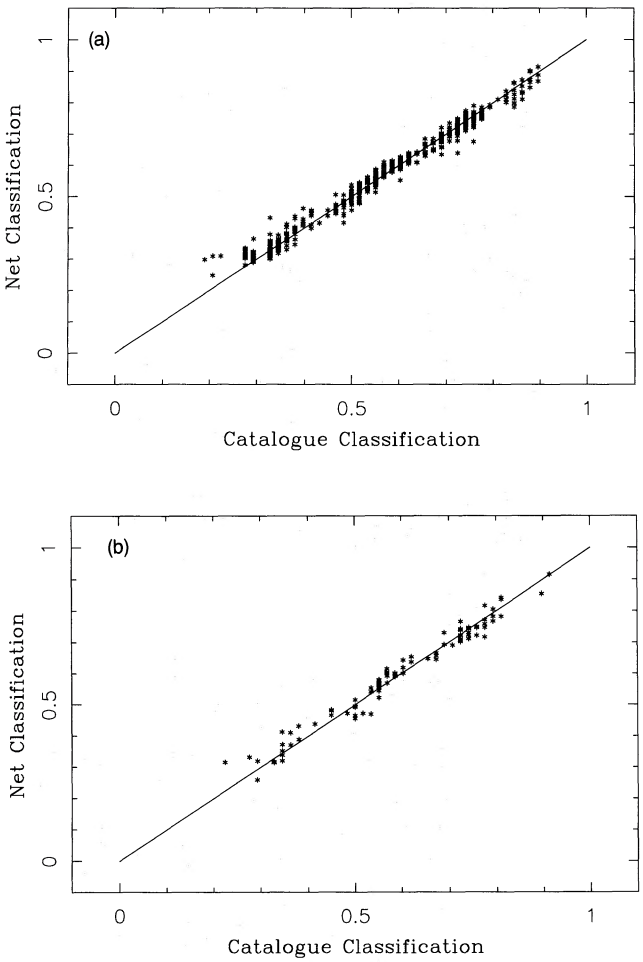


**Figure 4.** ANN results in the complete spectra case for typical training (a) and testing (b) sessions. The horizontal axes are the Houk classifications and the vertical axes are the ANN classifications. Unit-slope lines are overplotted.

$\sigma_{68}$ would be equal, but in general the distributions are not normal and have an excess of values very far from the mean. This statistic is added to aid in the discussion of those cases. While Table 2 gives the detailed results for each plate, Table 3 presents the summary statistics for each of the three cases. The first column in Table 3 identifies the cases (c+l = continuum plus line, c = continuum only, l = line only). The second to seventh columns are the same as those of Table 2 except that the mean of the absolute values of the systematic offsets and the mean of the classification uncertainties are reported.

Table 2 demonstrates that there is little difference between the plates, in terms of either the systematics or the rms uncertainties. The plate offsets vary from −0.7 to +1.1 spectral subtypes, while the uncertainties vary from 1.0 to 2.2 spectral subtypes. The mean systematic classification offset is ≈ 0.6 spectral subtypes and the mean rms uncertainty is ≈ 1.7 spectral subtypes, while 68 per cent of the objects are classified within ≈ 1.4 spectral subtypes.

Fig. 4 (and the statistics of Tables 2 and 3) also shows that no stars are grossly misclassified. Approximately 5 per cent of the spectra are still contaminated with overlapping stars, however, and this probably degrades the quality of the ANN results. We see systematic errors of ≈ 2 spectral subtypes

**Table 2.** Full spectra classification statistics.

| | train | | | test | | |
|---|---|---|---|---|---|---|
| plate | offset | $\sigma$ | $\sigma 68$ | offset | $\sigma$ | $\sigma 68$ |
| 2280 | -0.63 | 1.42 | 1.21 | -0.56 | 1.00 | 0.95 |
| 2303 | 0.35 | 1.40 | 1.22 | 0.78 | 1.67 | 1.27 |
| 3316 | -0.74 | 1.33 | 1.11 | -0.64 | 2.23 | 1.66 |
| 3504 | -0.54 | 1.42 | 1.24 | -0.74 | 1.83 | 1.72 |
| 4825 | 1.08 | 1.39 | 1.26 | 0.91 | 1.98 | 1.36 |
| 16506 | -0.18 | 1.34 | 1.14 | -0.18 | 1.59 | 1.57 |

**Table 3.** Summary statistics.

| | train | | | test | | |
|---|---|---|---|---|---|---|
| plate | offset | $\sigma$ | $\sigma 68$ | offset | $\sigma$ | $\sigma 68$ |
| c+l | 0.59 | 1.38 | 1.20 | 0.64 | 1.72 | 1.42 |
| c | 1.29 | 2.74 | 2.33 | 1.46 | 2.82 | 2.18 |
| l | 0.19 | 1.39 | 1.18 | 0.24 | 2.15 | 1.72 |

at the blue and/or red ends for some of the training and testing sets. The systematics for the bluest stars (B3-B5) are the result of the subtle features which cause the early B stars to differ from the later B stars, and the lack of examples in this temperature region. The systematics for the reddest stars (M3-M4) are the result of the diminishing amount of distinguishing spectral information available in blue spectra of very late-type stars, as well as the limited number of examples of these objects.

### 4.2 Line-only case

The line-only case was studied because human classifiers generally claim that they only pay attention to the line information. Additionally, while the complete spectra case provided better results, the line-only case has the potential to be less sensitive to reddening, plate emulsion sensitivity non-linearities and inter-plate differences.

Table 3 summarizes the results for the line-only ANNs. Here it can be seen that the line-only case has a smaller overall systematic offset but a slightly higher classification uncertainty ($\sigma = 2.2$ instead of 1.7, seen in the complete case). The lower overall systematic offset for the line-only case is most likely due to the sensitivity of the stellar continua shapes to reddening and systematic plate emulsion differences. Comparison of the complete-case classifications with the line-only classifications may eventually give reddening values for each star, once the plate emulsion differences are calibrated. The cause of the higher classification uncertainty in this case when compared to the complete spectra case is due to the $\approx 5$ per cent of the classifications that are very far removed from their nominal catalogue values, again leading to $\sigma > \sigma_{68}$. All stars that deviated grossly in the line-only case (we did not observe this effect in the other cases) displayed a large amount of noise somewhere in their spectra, generally in the blue. A further $\approx 10$ per cent of the stars had considerable noise in the blue and were still properly classified. This highlights the degree to which the ANNs were generally robust to noise.

For the line-only case the mean systematic classification offset is $\approx 0.2$ spectral subtypes and the uncertainty is $\approx 2.2$ spectral subtypes, while 68 per cent of the objects are classified within $\approx 1.7$ spectral subtypes.

### 4.3 Continuum-only case

The continuum-only case was studied because of its similarity to broad-band photometry. It is not our intention here to use our continuum-only data to compete with broad-band colours, but we wish to demonstrate that complementary information is contained in the continuum and in the lines. Broad-band colours can give very accurate temperatures (e.g. Böhm-Vitense 1989) if the luminosity class and the approximate metallicity of a star are known, but the broad-band colour distribution at a given temperature is generally bimodal for a mix of luminosity classes, even at constant metallicity. Thus broad-band colours cannot replace good quality spectra. Like the complete spectra case, the continuum spectra were normalized by setting the maximum intensity value to 1, and the minimum to 0.

Table 3 summarizes the results for the continuum-only ANNs. It can be seen here that the mean systematic classification offset is $\approx 1.5$ spectral subtypes and the uncertainty is $\approx 2.8$ spectral subtypes, while 68 per cent of the objects are classified within $\approx 2.2$ spectral subtypes. The correlation between continuum shape and spectral type is not surprising given the well-known correlation between stellar colour and temperature. The scatter is greater than that of the line case, however, indicating the effects of reddening, plate non-linearity and plate differences, as well as the small mixture of luminosity classes at a given temperature. No grossly deviant classifications exist, since noise is no longer significant for any of these objects. Systematic curvatures or offsets were present in most of the catalogue versus ANN classifications, however, and therefore $\sigma > \sigma_{68}$.

The continuum information and line information are complementary in the statistical sense, and the quality of classifications can be added in quadrature to get the uncertainties in the complete case, i.e.

$$\sigma^{-2}_{\text{expected}} = \sigma^{-2}_{\text{line}} + \sigma^{-2}_{\text{cont}} = 2.15^{-2} + 2.82^{-2} = 1.71,$$

essentially equal to the measured value of 1.72.

## 5 CONCLUSIONS

We have presented initial results in the automated determination of stellar spectral types using 575 stars scanned from six objective prism plates at the APM facility. Our automated techniques are based on a simple back-propagation artificial neural network which is trained to duplicate the classifications of Houk. We find that even with the limited data set now in hand we can determine high-quality temperature classifications for spectral subtypes from B3 to M4, though we do not yet have the ability to determine the luminosity classifications. We classify spectra to within 2.8 spectral subtypes based on continuum information alone, to within 2.2 spectral subtypes based on line information alone, and to within 1.7 spectral subtypes from the complete spectra.

Future work will be based on more plate scans so that we can build expert systems which have more adequate representations of nearly all spectral types and luminosities over a range of metallicities. We shall extend the ANN techniques to the determination of luminosity classes as well, and then

most likely use classical techniques to study stellar abundances. We also intend a more rigorous study of the effects of noise, resolution and wavelength calibration errors on ANN output.

In future papers we shall also explore alternative classification techniques, including unsupervised ANNs. The supervised networks presented here serve as error minimization algorithms, replicating the choices and decision patterns of their trainer. Unsupervised ANNs, however, are capable of deciding on an appropriate classification scheme without benefit of a trainer. Such unsupervised ANNs can learn to group or cluster objects into bins, usually by determining the Euclidean (or any other appropriate metric) distance between individual vectors. These ANNs operate much like classical vector analysis strategies, and we are exploring them since they may allow us to identify previously unrecognized stellar classifications.

Finally, we hope to extend the most useful techniques to fainter apparent magnitudes with the acquisition of deeper plate material and its calibration by Houk.

## ACKNOWLEDGMENTS

It is a pleasure to thank Nancy Houk for her considerable efforts in helping us begin this programme and for loaning us the plate material. We had assistance with numerical techniques from Geraint Lewis and Derek Richardson. We thank Ofer Lahav for very useful discussions and comments on the manuscript.

## REFERENCES

Aitken R.G., 1935, Binary Stars, 2nd edn. McGraw-Hill, London, p. 273
Böhm-Vitense E., 1989, Introduction to Stellar Astrophysics, Vol. 2. Cambridge Univ. Press, Cambridge, p. 13
Cannon A.J., Pickering E.C., 1918-1924, The Henry Draper Catalogue. Harvard Ann., 91-99 (HD)
Cardaliaguet P., Euvrard G., 1992, Neural Networks, 5, 207
de Villiers J., Barnard E., 1993, IEEE Trans. on Neural Networks, 4, 136
Eaton H.A.C., Olivier T.L., 1992, Neural Networks, 5, 283
Eggen O.J., Lynden-Bell D., Sandage A., 1962, ApJ, 136, 748
Gulati et al., 1994, ApJ, in press
Hertzsprung E., 1911, Potsdam Publ., 63, 1
Hewett P.C., Irwin M.J., Bunclark P.S., Bridgeland M.T., Kibblewhite E.J., He X.T., Smith M.G., 1985, MNRAS, 213, 971
Hopfield J.J., Tank D.W., 1986, Sci, 233, 625
Houk N., 1978, University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars, Vol. 2
Houk N., 1982, University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars, Vol. 3
Houk N., Cowley A.P., 1975, University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars, Vol. 1
Houk N., Smith-Moore M., 1988, University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars, Vol. 4
Jacoby et al., 1984, ApJS, 56, 257
Johnson H.L., Morgan W.W., 1953, ApJ, 117, 313
Jones D.H.P., 1966, R. Obs. Bull., 126, 219
Kibblewhite E.J., Bridgeland M.T., Bunclark P.S., Irwin M.J., 1984, in Klinglesmith D.A., ed., NASA-2317, Astronomical Microdensitometry Conference. NASA, Washington D.C., p. 277
Kurtz M.J., 1983, in Garrison B.F., ed., The MK Process and Stellar Classification. David Dunlop Observatory, Toronto, p. 136
McCullogh W.S., Pitts W.H., 1943, Bull. Math. Biophys., 5, 115
Merkle F., 1988, Messenger, 52, 5
Mihalas D., Binney J., 1981, Galactic Astronomy: Structure and Kinematics. W.H. Freeman & Co., San Fransisco, p. 111
Morgan W.W., Keenan P.C., Kellman E., 1943, An Atlas of Stellar Spectra. University of Chicago Press, Chicago
Mpitsos G.J., Burton R.M., 1992, Neural Networks, 5, 605
Odewahn E.B., Stockwell R.L., Pennington R.M., Humphreys R.M., Zumach W., 1991, AJ, 103, 318
Parker D.B., 1985, Learning-logic, Report TR-47. MIT Press, Cambridge, MA
Popper D.M., 1980, ARA&A, 18, 115
Richard M.D., Lippmann R.P., 1991, Neural Comput., 3, 461
Roman N.G., 1954, AJ, 59, 307
Rumelhart D.E., Hinton G.E., Williams R.J., 1986, in McClelland J.L., Rumelhart D.E., and the PDP Research Group, eds, Parallel Distributed Processing, Vol. 1. MIT Press, Cambridge, MA
Russell H.N., 1912, ApJ, 35, 315
Russell H.N., 1914, Nat, 93, 227
Sandler D.G., Parrett T.K., Palmer D.A., Fugate R.Q., Wild W.J., 1991, Nat, 351, 300
Silva & Cornell, 1992, ApJS, 81, 865
Storrie-Lombardi M., Lahav O., Sodré L., Storrie-Lombardi L., 1992, MNRAS, 259, 8P
Tukey J. W., 1971, Exploratory Data Analysis. Addison-Wesley, Reading
Werbos P.J., 1974, Ph.D. Thesis, Harvard University
Widrow B., Winter R., 1988, Computer, 21, 25
Wizinowich P., Lloyd-Hart M., McLeod B., Colucci D., Hulburd B., Snadler D., 1991, Proc. SPIE, 1542

## NOTE ADDED IN PROOF

After submitting this paper we became aware of the work of Gulati et al. (1994). They have also classified stellar spectra on the MK system with artificial neural networks, using spectral libraries (Jacoby et al. 1984; Silva & Cornell 1992) as training and testing sets. We refer the interested reader to their work.

This paper has been produced using the Blackwell Scientific Publications LaTeX style file.